Is Time Ripe for Fabric on a Chip?

Itamar Elhanany, University of Tennessee
Kurt Busch, Mindspeed Technologies

FoCs could play a key role in shaping the next-generation networking infrastructure.

At the heart of any switch or router is a packet-switching fabric subsystem. Broadly speaking, the switch fabric facilitates the exchange of data packets from the input ports to the output ports. Fabrics range in size from hand-sized, four-port boxes to rack-mounted chassis that require forklifts to reposition. Port density and data rates strongly dictate fabric complexity.

Advances in point-to-point communication technologies necessitate continuous scalability of fabric subsystems. Implementations of these subsystems range from shared-memory designs to fully distributed, multistage architectures. Regardless of the particular design, however, most high-capacity fabrics buffer incoming packets at the ingress modules until a scheduling mechanism signals them to traverse the switch core.

In typical implementations, buffering resides at the line cards, while the switch cards contain very little memory. In addition to payload data, control information also flows to and from the line cards and the fabric, ensuring perpetual switch operation.

Today's backbone Internet routers support traffic capacities on the order of multiple terabits per second, while systems built for metropolitan area networks (MANs) typically carry hundreds of gigabits per second. In both cases, the switch fabrics that power these systems comprise a myriad of chips that span multiple boards and perform key functions including address lookup, packet prioritization, queue arbitration, scheduling, and switching and buffering.

MARKET TRENDS

Over the past 20 years, switch fabric development has followed a familiar pattern. Shortly following publication of a communications standard, researchers and vendors create multichip solutions to support it. Advances in silicon technology subsequently lead to a single-chip solution that often provides equal or greater utility at a much lower price. This in turn accelerates adoption of the standard. Once the standard has widespread acceptance, a new standard emerges and the cycle continues.

For example, Ethernet, the most commonly exploited networking standard, has progressed from 10 megabits per second through 100 Mbps to 1,000 Mbps, in each case with initial multichip solutions followed by single-chip solutions and extensive market adoption. Multichip fabrics capable of tens of Gbps are now readily available, and researchers are currently exploring ways to migrate these solutions to single-chip configurations.

During the late 1990s, many believed that rapid Internet growth would require a significant upgrade in switching infrastructure as often as every 18 months. Although user traffic is roughly doubling each year, the pragmatic requirements of backbone switches and routers are more modest. Nevertheless, the numerous components in switch fabrics, which drive such large systems, make their design, testing, and maintenance highly complex.

Thus, a key question is whether time is ripe to alleviate some of these difficulties by introducing *fabric on a chip* (FoC)—a single-chip, high-capability switch fabric solution.

CONSOLIDATING SWITCHING FUNCTIONS

Moving from expensive multichip fabrics to low-cost single-chip ones presents design challenges with respect to the

- processing power required to support the decision-making process involved;
- amount of on-chip memory needed to store "in-flight" packets—those that have entered the switch but have yet to depart through the designated output ports; and
- packaging and I/O bandwidth considerations that inherently determine the upper limit of switching capacity.

Leveraging recent very large-scale integration advances, an FoC would consolidate as many core switching functions as possible on a single chip.

Achieving this high level of integration arguably could lead to the straightforward realization of much larger systems. Moreover, the resulting designs would consume significantly lower resources compared to the traditional approach as well as facilitate interoperability and manageability.

Given current I/O bandwidth limitations, near-term FoCs clearly would not be feasible for core/backbone routers. Rather, the target application space would be equipment used in MANs, high-end local area networks, and storage area networks that carry up to hundreds of Gbps.

MEMORY MANAGEMENT BOTTLENECK

Output-queued (OQ) switches offer several highly desirable performance characteristics including

- minimal average packet delay,
- controllable quality-of-service (QoS) provisioning, and
- work conservation under any admissible traffic conditions.

However, the memory bandwidth requirements of such switches is $O(NR)$, where N denotes the number of ports and R each port's data rate. The latter is derived from the need to be able to accept (write) up to N arriving packets and, simultaneously, transmit (read) up to N departing packets. This requirement significantly limits the scalability of OQ switches with respect to aggregate switching capacity.

Parallel shared memory (PSM) switches mimic the desirable attributes of OQ switches while significantly reducing their memory bandwidth requirements. The underlying assumption is that multiple memories, each having relatively low bandwidth, can collectively emulate an OQ switch. To achieve this goal, a PSM switch contains a memory management algorithm that assigns a nonconflicting memory unit to each arriving packet.

Despite PSM's great potential, the complexity of the memory management algorithms developed thus far has limited its scalability. Recent research suggests that alternative realizations of the algorithm, such as those exploiting a time-space tradeoff, might create the much-needed breakthrough that will ultimately enable large FoCs.

A primary effort in this context focuses on utilizing pipeline structures that introduce a fixed delay in the packet-to-memory assignment process, yet can support the required high switching throughput (in terms of packets per second). Enhancing these techniques with the ability to support QoS provisioning is a fundamental research challenge.

Once the memory management bottleneck is resolved, the I/O bandwidth that the device pads support is bound to determine FoC capacity. The ongoing evolution in integrated high-speed serializer-deserializer (SerDes) technology will help facilitate FoC solutions. Current high-speed serial links operate at 6.25 Gbps, and 10-Gbps cores will soon be available.

The future holds great promise for FoC technology, with the potential to play a key role in shaping the next-generation networking infrastructure. In view of the intricate engineering challenges lying ahead, FoCs will become an active subject of research, spanning both industry and academia.

Itamar Elhanany *is an assistant professor in the Department of Electrical and Computer Engineering at the University of Tennessee. Contact him at itamar@ieee.org.*

Kurt Busch *is director of business development at Mindspeed Technologies, a provider of semiconductor networking solutions based in Newport Beach, California. Contact him at kurt.busch@mindspeed.com.*

Editor: Richard G. Mathieu, Dept. of Computer Information Systems and Management Science, College of Business, James Madison Univ., Harrisonburg, VA; mathierg@jmu.edu